

# Lower Complexity Bounds for Lifted Inference

Manfred Jaeger

Department of Computer Science, Aalborg University, Denmark  
jaeger@cs.aau.dk

March 2, 2013

## Abstract

One of the big challenges in the development of probabilistic relational (or probabilistic logical) modeling and learning frameworks is the design of inference techniques that operate on the level of the abstract model representation language, rather than on the level of ground, propositional instances of the model. Numerous approaches for such “lifted inference” techniques have been proposed. While it has been demonstrated that these techniques will lead to significantly more efficient inference on some specific models, there are only very recent and still quite restricted results that show the feasibility of lifted inference on certain syntactically defined classes of models. Lower complexity bounds that imply some limitations for the feasibility of lifted inference on more expressive model classes were established early on in (Jaeger 2000). However, it is not immediate that these results also apply to the type of modeling languages that currently receive the most attention, i.e., weighted, quantifier-free formulas. In this paper we extend these earlier results, and show that under the assumption that  $\text{NETIME} \neq \text{ETIME}$ , there is no polynomial lifted inference algorithm for knowledge bases of weighted, quantifier- and function-free formulas. Further strengthening earlier results, this is also shown to hold for approximate inference, and for knowledge bases not containing the equality predicate.

## 1 Introduction

Probabilistic logic models (a.k.a. probabilistic or statistic relational models) provide high-level representation languages for probabilistic models of structured data. One broad distinction one can make for the very many different types of models that have been proposed is between what one may call *process-oriented* and *constraint-based* models. In the former, the model essentially represents a generative stochastic process for a relational structure. It is defined in terms of prior and conditional probabilities, and its semantics can be given in terms of a directed graphical model, e.g., [1, 18, 21, 17, 9, 6, 13, 15, 27] – and many others. In the latter, the model defines a set of soft

constraints on a distribution, and the semantics can be given in terms of an undirected graphical model [24, 20].

The problem of finding “lifted” inference techniques for probabilistic logic models has received a lot of attention for nearly 10 years now [19, 2, 16, 14, 10, 7, 26, 25, 5]. Apart from methods for exact lifted inference, also approximate lifted techniques have been investigated [22, 3, 23]. In most of these works it is demonstrated for some particular models that lifted inference techniques provide a significant improvement in terms of how inference complexity scales as a function of the size of the model domain.

On the other hand, [8] has shown that under certain assumptions on the expressivity of the modeling language, probabilistic inference is not polynomial in the domain size, thereby demonstrating some inherent limitations in terms of worst-case complexity for the goals of lifted inference. The results of [8] essentially assume a process-oriented modeling framework, and the expressivity requirements amount to a probabilistic version of full first-order predicate logic. Since much recent work considers constraint-based frameworks, and, more significantly, within these frameworks focuses on fragments without full first-order expressivity, it is not clear to what extent these earlier intractability results are applicable to these ongoing efforts. In particular, much current work is devoted to models defined by constraints expressed by quantifier- and function-free formulas, which do not fulfill the requirements of [8]. In fact, for one limited sub-class of such models (defined by the condition that formulas contain at most two variables), [25] has recently proposed a lifted inference technique that guarantees polynomial complexity. This appears to be the first result that establishes scalability of lifted inference for a whole model class defined by syntactic restrictions. Between this positive result, and the earlier negative results, the theoretical complexity boundaries for lifted inference are unknown.

In this paper we refine the complexity map for lifted inference. Extending the general approach taken in [8], we establish intractability results also for constraint-based modeling languages limited to quantifier- and function-free formulas. In a sharp contrast with [8], where a “trivial” constant-time approximate inference method was described, we show that our lower complexity bounds also hold for approximate inference. Further sharpening earlier results, we finally establish that the lower complexity bounds also hold for models not using the equality predicate, which in [8] was conjectured to be the key source of inherent complexity.

In the following section we briefly review the inference problem for constraint-based probabilistic logic models in terms of weighted model counting. Section 3 reviews classic results relating first-order logic models to the complexity class `NETIME`. Section 4 contains our main results, and Section 5 discusses some notable differences that emerge between the results for process-oriented and for constraint-based models.

## 2 Weighted Model Counting

Similarly as [26] and [7] we assume the following framework: a model, or knowledge base, is given by a set of weighted formulas:

$$KB : \begin{array}{ll} \phi_1(\mathbf{v}_1) & : w_1 \\ \phi_2(\mathbf{v}_2) & : w_2 \\ \dots & \dots \\ \phi_n(\mathbf{v}_n) & : w_n \end{array} \quad (1)$$

where the  $\phi_i$  are formulas in first-order predicate logic,  $w_i \in \mathbb{R}$  are non-negative *weights*, and  $\mathbf{v}_i = (v_{i,1}, \dots, v_{i,k_i})$  are the free variables of  $\phi_i$ . The case  $k_i = 0$ , i.e.,  $\phi_i$  is a sentence without free variables, is also permitted. The  $\phi_i$  use a given signature  $S$  of relation-, function-, and constant symbols.

An interpretation (or possible world)  $(D, I)$  for  $S$  consists of a domain  $D$ , and an interpretation function  $I$  that maps the symbols in  $S$  to functions, relations and elements on  $D$ . For a tuple  $\mathbf{d} \in D^{k_i}$  then the truth value  $\phi_i(\mathbf{d}/\mathbf{v}_i)$  is defined, and we write  $(D, I) \models \phi_i(\mathbf{d})$ , or simpler  $I \models \phi_i(\mathbf{d})$  if  $\phi_i(\mathbf{d}/\mathbf{v}_i)$  is true in  $(D, I)$ . We use  $\mathcal{I}(D, S)$  to denote the set of all interpretations for the signature  $S$  over the domain  $D$ .

In this paper we are only concerned with finite domains, and assume without loss of generality that  $D = D_n := \{1, \dots, n\}$  for some  $n \in \mathbb{N}$ .

For  $I \in \mathcal{I}(D_n, S)$  let  $\#(i, I)$  denote the number of elements  $\mathbf{d}$  in  $D^{k_i}$  for which  $I \models \phi_i(\mathbf{d})$ . The *weight* of  $I$  then is

$$w_n^{KB}(I) := \prod_{i=1}^N w_i^{\#(i, I)},$$

where  $0^0 = 1$ . The probability of  $I$  is

$$P_n^{KB}(I) = w_n^{KB}(I)/Z$$

where  $Z$  is the normalizing constant (partition function)

$$Z = \sum_{I' \in \mathcal{I}(D_n, S)} w_n^{KB}(I'). \quad (2)$$

For a first-order sentence  $\phi$  and  $n \in \mathbb{N}$  then

$$P_n^{KB}(\phi) := P(\{I \in \mathcal{I}(D_n, S) \mid I \models \phi\})$$

is the probability of  $\phi$  in  $\mathcal{I}(D_n, S)$ . An *inference problem*  $PI(KB, n, \phi, \psi)$  is given by a knowledge base  $KB$ , a domainsize  $n \in \mathbb{N}$ , and two first-order sentences  $\phi, \psi$ . The solution to the inference problem is the conditional probability  $P_n^{KB}(\phi \mid \psi)$ .

A *class* of inference problems is defined by allowing arguments  $KB$ ,  $\phi$ , and  $\psi$  only from some restricted classes  $\mathcal{KB}$ ,  $\mathcal{Q}$  (the query class), and  $\mathcal{E}$  (the evidence class), respectively. In this paper we will only be concerned with the cases where  $\mathcal{Q}$  consists of all ground atoms, denoted  $\mathcal{AT}$ , and  $\mathcal{E}$  is empty, i.e., we are only considering inference without evidence. Classes  $\mathcal{KB}$  are defined by various syntactic restrictions on the

formulas  $\phi_i$  in the knowledge base. In this paper, we consider the following fragments of first-order logic (FOL): relational FOL (RFOL), i.e. FOL without function and constant symbols; 0-RFOL, which is the quantifier-free fragment of RFOL, and 0-RFOL $^\neq$ , which is 0-RFOL without the equality relation.

An algorithm solves a class  $PI(\mathcal{KB}, \mathbb{N}, \mathcal{Q}, \mathcal{E})$ , if it computes  $P_n^{KB}(\phi \mid \psi)$  for all instances  $PI(KB, n, \phi, \psi)$  in the class. An algorithm  $\epsilon$ -approximately solves  $PI(\mathcal{KB}, \mathbb{N}, \mathcal{Q}, \mathcal{E})$ , if for any  $PI(KB, n, \phi, \psi)$  in the class it returns a number  $p \in [P_n^{KB}(\phi \mid \psi) - \epsilon, P_n^{KB}(\phi \mid \psi) + \epsilon]$ . An algorithm that solves  $PI(\mathcal{KB}, \mathbb{N}, \mathcal{Q}, \mathcal{E})$  is *polynomial in the domainsize*, if for fixed  $KB, \phi, \psi$  the computation of  $PI(KB, n, \phi, \psi)$  is polynomial in  $n$ .

### 3 Spectra and Complexity

The following definition introduces the central concept for our analysis.

**Definition 3.1** *Let  $\psi$  be a sentence in first-order logic. The spectrum of  $\psi$  is the set of integers  $n \in \mathbb{N}$  for which  $\psi$  is satisfiable by an interpretation of size  $n$ .*

**Example 3.2** *Let  $\psi = \psi_1 \wedge \psi_2 \wedge \psi_3$ , where*

$$\begin{aligned}\psi_1 &\equiv \forall x, y \ u(x, y) \Leftrightarrow u(y, x) \\ \psi_2 &\equiv \forall x \exists y \ u(x, y) \\ \psi_3 &\equiv \forall x, y, y' \ (u(x, y) \wedge u(x, y') \Rightarrow y = y')\end{aligned}$$

*$\psi$  expresses that the binary relation  $u$  defines an undirected graph ( $\psi_1$ ) in which every node is connected to exactly one other node ( $\psi_2, \psi_3$ ). Thus,  $\psi$  describes a pairing relation that is satisfiable exactly over domains of even size:  $\text{spec}(\psi) = \{n \mid n \text{ even}\}$ .*

The complexity class ETIME consists of problems solvable in time  $O(2^{cn})$ , for some constant  $c$ . The corresponding nondeterministic class is NETIME. Note that these classes are distinct from the more commonly studied classes (N)EXPTIME, which are characterized by complexity bounds  $O(2^{n^c})$  [11]. For  $n \in \mathbb{N}$  let  $\text{bin}(n) \in \{0, 1\}^*$  denote the binary coding of  $n$ , and  $\text{un}(n) \in \{1\}^*$  the unary coding (i.e.,  $n$  is represented as a sequence of  $n$  1s). A set  $S \subseteq \mathbb{N}$  is in (N)ETIME, iff  $\{\text{bin}(n) \mid n \in S\}$  is in (N)ETIME, which also is equivalent to  $\{\text{un}(n) \mid n \in S\}$  being in (N)PTIME.

Like [8], we use the following connection between spectra and NETIME as the key tool for our complexity analysis.

**Theorem 3.3** [12] *A set  $A \subseteq \mathbb{N}$  is in NETIME, iff  $A$  is the spectrum of a sentence  $\phi \in \text{RFOL}$ .*

**Corollary 3.4** *If  $\text{NETIME} \neq \text{ETIME}$ , then there exists a first-order sentence  $\phi$ , such that  $\{\text{un}(n) \mid n \in \text{spec}(\phi)\}$  is not recognized in deterministic polynomial time.*

Thus, by reducing the spectra-recognition problem to a class of inference problems  $PI(\mathcal{KB}, \mathbb{N}, \mathcal{Q}, \mathcal{E})$ , one establishes that the latter is not polynomial in the domainsize (under the assumption  $\text{ETIME} \neq \text{NETIME}$ ).

## 4 Complexity Results

This section contains our complexity results. We begin with a result for knowledge bases using full RFOL. This is rather straightforward, and (for exact inference) already implied by the results of [8]. We then proceed to extend this base result to 0-RFOL and 0-RFOL $^\neq$ .

**Theorem 4.1** *If  $NETIME \neq ETIME$ , then there does not exist an algorithm that 0.25-approximately solves  $PI(RFOL, \mathbb{N}, \mathcal{AT}, \emptyset)$  in time polynomial in the domain size.*

The proof of this theorem provides the general pattern also for subsequent proofs. It is therefore here given in full.

**Proof:** Let  $\phi$  be a sentence with a “hard” spectrum as given by Corollary 3.4. Let  $S$  be the relational signature of  $\phi$ . Let  $a()$  be a new relation symbol of arity zero (i.e.,  $a()$  represents a propositional variable). The first weighted formula in our knowledge base then is

$$\neg(\phi \leftrightarrow a()) \quad : \quad 0 \quad (3)$$

We now already have that  $P_n^{KB}(a()) > 0$  iff there exists  $I \in \mathcal{I}(D_n, S)$  with  $I \models \phi$ , i.e., iff  $n \in spec(\phi)$ . This already reduces the decision problem for  $spec(\phi)$  to solving  $PI(KB, n, a(), \emptyset)$  exactly. However, from the 0-1 laws of first-order logic [4], it follows that for our current  $KB$ :  $P_n^{KB}(a()) \rightarrow_{n \rightarrow \infty} 0$ . Thus, for every  $\epsilon > 0$  we could define an  $\epsilon$ -approximate constant-time inference algorithm by returning 0 for all sufficiently large  $n$ .

In order to obtain our result for approximate inference, we will now ensure that for all  $n \in spec(\phi)$  the probability  $P_n^{KB}(a())$  is greater than 0.5. We do this essentially by calibrating the normalization constant  $Z$  in (2). For this we introduce another new relation  $b()$ , and add to  $KB$ :

$$\neg((\bigwedge_{R \in S} \forall x \neg R(x)) \leftrightarrow b()) \quad : \quad 0 \quad (4)$$

Thus, for every  $n$  there is exactly one interpretation  $I \in \mathcal{I}(D_n, S)$  with nonzero weight in which  $b()$  is true (the one in which all relations have empty interpretations). Finally, we give zero weight to all interpretations except those in which  $a()$  or  $b()$  is true:

$$\neg(a() \vee b()) \quad : \quad 0 \quad (5)$$

Let  $KB$  consist of (3),(4),(5). Every  $I \in \mathcal{I}(D_n, S)$  then has weight 0 if it satisfies one of the three formulas, and weight 1 otherwise. Consider the case  $n \notin spec(\phi)$ . Then, by (3)  $w_n^{KB}(a()) = 0$ . By (5) this then means that in all interpretations of nonzero weight  $b()$  must be true. By (4) there is exactly one such interpretation. Thus,  $Z$  in (2) is 1, and  $P_n^{KB}(a()) = 0/1 = 0$ .

If  $n \in spec(\phi)$ , then  $w_n^{KB}(a()) \geq 1$ , and  $Z = w_n^{KB}(a())$  (if the interpretation in which all  $R$  are empty also is a model of  $\phi$ ), or  $Z = w_n^{KB}(a()) + 1$  (otherwise). Thus,  $P_n^{KB}(a()) \geq 1/2$ . A 0.25-approximate inference algorithm for  $PI(KB, n, a(), \emptyset)$ , thus, would decide  $spec(\phi)$ .  $\square$

We now proceed towards our main result, which is going from RFOL to 0-RFOL. If we wanted to allow function and constant symbols in our knowledge base, then one could go to a quantifier-free fragment in a quite straightforward manner using Skolemization. Since satisfiability over a given domain is the same for a formula  $\phi$  and its quantifier-free Skolemized version  $\phi^{Skol}$ , the arguments of the proof of Theorem 4.1 would go through with little change. In order to accomplish the same using only the relational fragment 0-RFOL, we define the *relational Skolemization* of a formula. The idea is to replace function and constant symbols in the Skolemized version of a formula with relational representations. For example, the Skolemized version of  $\psi_2$  from Example 3.2 is

$$\psi_2^{Skol} \equiv \forall x \ u(x, f(x))$$

with a new function symbol  $f()$ . Introducing a relational encoding of  $f()$  leads to

$$\psi_2^{R-Skol} \equiv \forall x, y \ R^f(x, y) \rightarrow u(x, y)$$

with  $R^f$  a new relation symbol encoding  $f()$ . This translation must be accompanied by axioms that confine the possible interpretations of  $R^f$  to relations that encode functions.

Such relational encodings of functions are well established. However, there does not seem to be a standard account of this technique that serves our purpose. The following proposition, therefore, provides the relevant result in a form tailored for our needs.

**Proposition 4.2** *Let  $\phi(x) \in 0-FOL(S \cup S^F)$ , where  $S$  is a set of relation symbols, and  $S^F$  a set of function and constant symbols. Let  $S^+$  be a set of new relation symbols that for every  $k$ -ary  $f \in S^F$  contains a  $k+1$ -ary  $R^f$  (constant symbols are treated as 0-ary function symbols). Let  $\text{Func}$  be the set of sentences that for every  $f \in S^F$  contains*

$$\forall x \ y \ y' \ (R^f(x, y) \wedge R^f(x, y') \rightarrow y = y') \quad (6)$$

$$\forall x \exists y \ R^f(x, y) \quad (7)$$

*Then there exists a formula  $\phi^+(x, z) \in 0-RFOL(S \cup S^+)$ , such that the following are equivalent:*

- i** *there exists  $I \in \mathcal{I}(D_n, S \cup S^F)$  with  $I \models \forall x \phi(x)$*
- ii** *there exists  $I^+ \in \mathcal{I}(D_n, S \cup S^+)$  with  $I^+ \models \text{Func} \wedge \forall x z \ \phi^+(x, z)$*

If  $\phi^{Skol}$  is the Skolemization of a formula  $\phi \in \text{RFOL}$ , we then call  $\phi^{Skol+}$  the *relational Skolemization* of  $\phi$ , written  $\phi^{R-Skol}$ .

Our plan, now, is to prove the analogon of Theorem 4.1 for 0-RFOL by replacing  $\phi$  in (3) with  $\phi^{R-Skol}$ . However, this is not enough, since we also need to constrain the models of our knowledge base (more precisely: those models in which  $a()$  is true) to satisfy the axioms (6) and (7). This poses a problem, because (7) contains an existential quantifier, and so we cannot add this axiom directly as a constraint to a knowledge base restricted to 0-RFOL.

The solution to this problem is to approximate (7) with a weighted formula

$$a() \wedge R^f(x, y) \quad : \quad w \tag{8}$$

that rewards models of  $a()$  in which the existential quantifier of (7) is satisfied for many (all)  $x$ . We will no longer be able to ensure that  $w_n^{KB}(a()) = 0$  when  $n \notin \text{spec}(\phi)$ . However, by a suitable choice of  $w$ , and by a careful calibration of the weight of models of the alternative proposition  $b()$ , we still can ensure that  $w_n^{KB}(b()) \gg w_n^{KB}(a())$  when  $n \notin \text{spec}(\phi)$ , and  $w_n^{KB}(b()) \approx w_n^{KB}(a())$  when  $n \in \text{spec}(\phi)$ . This choice of  $w$ , however, now will have to be a function of  $n$ , i.e., we cannot reduce the decision problem of  $\text{spec}(\phi)$  to probabilistic inference for a fixed knowledge base  $KB$ , but to probabilistic inference for a parameterized knowledge base  $KB(w(n))$ . This is why the following theorem is a little weaker than the previous one in that it also includes a condition on the inference algorithm to be polynomial in the representation size of the weight parameters.

**Theorem 4.3** *If  $\text{NETIME} \neq \text{ETIME}$ , then there does not exist an algorithm that 0.25-approximately solves  $\text{PI}(0\text{-RFOL}, \mathbb{N}, \mathcal{AT}, \emptyset)$  in time polynomial both in the domain-size, and the representation size  $l := \sum_{i=1}^N \log(w_i)$  of the weight parameters.*

The full proof of the theorem is given in the appendix.

One may wonder how strong or surprising Theorem 4.3 really is in light of its extra polynomial runtime in  $l$  condition – especially since it has been emphasized that lifted inference procedures should only be expected to be polynomial in the domain size, but not in other parameters that characterize the complexity of  $KB$  [8, 25]. These remarks, however, have mostly been motivated by considerations of the logical complexity of  $KB$ , e.g. in terms of the number and complexity of its weighted formulas, or the size of the signature. The complexity in terms of numerical parameters, on the other hand, has not received much attention.

To better understand the nature of the polynomial in  $l$  condition, we consider inference algorithms that can be described as follows: to compute  $\text{PI}(KB, n, \phi, \psi)$  the algorithm performs a number of steps, where step  $i$  either consists of executing a constant time operation that does not depend on the numerical model parameters (e.g., a logical operation on formulas), or of an operation on numbers  $k, l \in V(i)$ , where  $V(i)$  is the set of all numerical variables (original weight parameters, intermediate computed results, ...) stored by the algorithm at step  $i$ . Further assume that the values of the weight parameters in  $KB$  only affect the numerical values of the variables in  $V(i)$ , but not the sequence of execution steps performed by the algorithm, and, in consequence, not the set of variables contained in  $V(i)$ . Also assume that an operation performed on  $k, l \in V(i)$  is linear in the representation size of  $k$  and  $l$  (as is the case for basic addition and multiplication operations). Then the representation size of the values in  $V(i)$ , and the execution time of a numerical computation step will always be a linear function of the size of the original  $w_i$  parameters, and the algorithm is polynomial (linear, in fact) in  $l$ . Since most existing lifted inference algorithms fit this description, it seems that the condition of being polynomial in  $l$  is not a severe limitation on the applicability of Theorem 4.3.

In a final strengthening of our results, we now move on to the fragment  $0\text{-RFOL}^\neq$ . The availability of the equality predicate for the formulas of  $KB$ , so far, has been an important prerequisite for our arguments, because Theorem 3.3 crucially depends on equality: spectra for formulas  $\phi \in \text{RFOL}^\neq$  are always of the form  $\mathbb{N} \setminus \{1, \dots, k\}$  for some  $k$ , and, thus, decidable in constant time. For this reason it was suggested in [8] that one should focus on logical fragments without equality when looking for model classes for which lifted inference scales polynomially in the domain size. As our final result shows, however, elimination of equality may not have such a large impact on complexity, after all.

**Theorem 4.4** *If  $\text{NETIME} \neq \text{ETIME}$ , then there does not exist an algorithm that 0.25-approximately solves  $\text{PI}(0\text{-RFOL}^\neq, \mathbb{N}, \mathcal{AT}, \emptyset)$  in time polynomial in  $n$  and the representation size  $l := \sum_{i=1}^N \log(w_i)$  of the weight parameters.*

This theorem is a generalization of Theorem 4.3, and, strictly speaking, makes 4.3 redundant. It is only for expository purposes, and greater transparency in the proof arguments, that we here develop these results in two steps.

The proof of Theorem 4.4 is a refinement of the proof of Theorem 4.3. In addition to approximating Skolem functions  $f$  with relations  $R^f$ , we now also approximate the equality predicate  $=$  with a binary relation  $E(\cdot, \cdot)$ . Similarly as we could not impose in  $0\text{-RFOL}$  hard constraints that ensure that  $R^f$  encodes a function, we also cannot constrain models to exactly interpret  $E$  as the equality relation. However, in analogy to (8) we can approximate true equality using the two weighted formulas

$$a() \wedge \neg E(x, x) \quad : \quad 0 \tag{9}$$

$$a() \wedge E(x, y) \quad : \quad 1/w \tag{10}$$

where  $w$  is a large weight. Any model that does not interpret  $E$  as the equality relation, then incurs a penalty of at least  $1/w$ .

## 5 Approximate Inference and Convergence

There are some notable differences with respect to approximate inference between the results we here obtained for weighted model counting, and the results of [8], where it was shown that due to convergence of query probabilities as  $n \rightarrow \infty$ , in theory a trivial constant time approximation algorithm exists: perform exact inference for all input domains up to a size  $n^*$ , and output the limit probability for all domains of size  $> n^*$ . This “algorithm”, however, has no practical use, since for a desired accuracy value  $\epsilon$  one first would have to determine a sufficiently high threshold value  $n^* \in \mathbb{N}$  to make the output indeed be an  $\epsilon$ -approximation.

Nevertheless, the difference between the existence of an impractical approximation algorithm on the one hand, and the non-existence of any approximation algorithm on the other hand, is just one consequence of a more fundamental difference: while in the models considered in [8] query probabilities  $P_n(a())$  converge to a limit, this is not the case for knowledge bases of weighted formulas – even under the restriction to  $0\text{-RFOL}$ : in the proofs of Section 4 we have constructed knowledge bases  $KB$ , such that



$P_n^{KB}(a())$  oscillates between zero and values  $> 1/2$  as  $n$  oscillates between  $\text{spec}(\phi)$  and its complement. Note that the construction of knowledge bases with this behavior does not require formulas  $\phi$  with a hard spectrum as in Corollary 3.4, and is not contingent on  $\text{NETIME} \neq \text{ETIME}$ . Already a knowledge base as constructed in the proof of Theorem 4.1 with  $\phi$  replaced by  $\psi$  of Example 3.2 will show this behavior.

The reason behind these different convergence properties lies in a somewhat different role that conditioning on evidence plays in process-oriented and constraint-based models: in the former, a conditional probability  $P_n^M(a() \mid b())$  defined by a model  $M$  can, in general, not be defined as an unconditional probability  $P_n^{M'}(a())$  in a modified model  $M'$ . For constraint knowledge bases  $KB$ , on the other hand, one can just add to  $KB$  the hard constraint  $\neg b() : 0$  to obtain  $KB'$  with  $P_n^{KB'} = P_n^{KB} \mid b()$ . Thus, there are here no fundamental differences between conditional and unconditional probabilistic queries. For procedural models, on the other hand, this difference is instrumental for the convergence of query probabilities: such a convergence only is guaranteed for unconditional queries, and can easily fail for conditional ones.

## 6 Conclusion

We have shown that for currently quite popular relational probabilistic models consisting of collections of weighted, quantifier- and function-free formulas there is likely to be no general polynomial lifted inference method (contingent on  $\text{NETIME} \neq \text{ETIME}$ ). Somewhat surprisingly, this even holds for approximate inference. Between this negative result, and the positive result of [25], there still could be a lot of room for identifying tractable fragments by restricting 0-RFOL further via limits on the number of variables, or the richness of the signature  $S$ .

## A Proofs

### Proof of Proposition 4.2:

We begin by defining the *term-depth* of a term  $t$  in the signature  $S^F$  as the maximal nesting depth of function symbols in  $t$ . Precisely, we define inductively: if  $t \equiv x$ , then  $t$  has term depth 0. If  $t \equiv f()$  (a constant), or  $t = f(x_1, \dots, x_k)$  (a function term with only variables as arguments), then  $t$  has term depth 1. If  $t = f(t_1, \dots, t_k)$ , then the term depth of  $t$  is one plus the maximal term depth of the  $t_i$ .

The term depth of a formula  $\phi(x)$  is the maximal term depth of the terms it contains.

We now show that every formula  $\phi(x)$  of term depth  $l$  can be transformed into a formula  $\phi^{l-1}(x, z)$  of term depth  $l-1$  in  $0\text{-FOL}(S \cup S^F \cup S^+)$ , such that the statement for  $\phi^+$  of the proposition holds for  $\phi^{l-1}$  (but with  $S \cup S^F \cup S^+$  instead of  $S \cup S^+$  in **ii**). The proposition then follows by defining  $\phi^+$  as the result of iteratively applying  $l$  such transformations to  $\phi$ . Since the term depth of the resulting  $\phi^+$  is zero, then actually  $\phi^+(x, z) \in 0\text{-RFOL}(S \cup S^+)$ .

Let  $\{f_i(x_i) \mid i = 1, \dots, r\}$  be the set of all distinct terms (including sub-terms) of

depth 1 appearing in  $\phi(\mathbf{x})$ . Let  $z_1, \dots, z_r$  be new variables. Define  $\phi^{l-1}(\mathbf{x}, \mathbf{z})$  as

$$\bigwedge_{i=1}^r R^{f_i}(\mathbf{x}_i, z_i) \rightarrow \phi(\mathbf{x})[z_1/f_1(\mathbf{x}_1), z_r/\dots, f_r(\mathbf{x}_r)]$$

To now show **i** $\Rightarrow$ **ii** let  $I \in \mathcal{I}(n, S \cup S^F)$  with  $I \models \forall \mathbf{x} \phi(\mathbf{x})$ . Define  $I^+ \in \mathcal{I}(n, S \cup S^F \cup S^+)$  as the expansion of  $I$  in which each  $R^f \in S^+$  is interpreted as the relational representation of  $f$ , i.e.,  $I^+ \models R^f(\mathbf{d}, e)$  iff  $I \models f(\mathbf{d}) = e$ . Clearly,  $I^+ \models \text{Func}$ . Furthermore, the following are equivalent:

$$\begin{aligned} I &\models \forall \mathbf{x} \phi(\mathbf{x}) \\ I &\models \forall \mathbf{x} \mathbf{z} \bigwedge_{i=1}^r f_i(\mathbf{x}_i) = z_i \\ &\quad \rightarrow \phi(\mathbf{x})[z_1/f_1(\mathbf{x}_1), z_r/\dots, f_r(\mathbf{x}_r)] \\ I^+ &\models \forall \mathbf{x} \mathbf{z} \bigwedge_{i=1}^r R^{f_i}(\mathbf{x}_i, z_i) \\ &\quad \rightarrow \phi(\mathbf{x})[z_1/f_1(\mathbf{x}_1), \dots, z_r/f_r(\mathbf{x}_r)] \end{aligned}$$

For **ii** $\Rightarrow$ **i** let  $I^+$  as in **ii** be given. Since  $I^+ \models \text{Func}$ , we can turn  $I^+$  into an interpretation for  $S \cup S^F$  by defining  $f(\mathbf{d})$  as the unique  $e$  for which  $R^f(\mathbf{d}, e)$  holds in  $I^+$ . Then, by the same equivalences as above,  $I^+ \models \forall \mathbf{x} \mathbf{z} \phi^+(\mathbf{x}, \mathbf{z})$  implies  $I \models \forall \mathbf{x} \phi(\mathbf{x})$ .  $\square$

**Proof of Theorem 4.3:** Let  $\phi \in \text{RFOL}$  as given by Corollary 3.4, and  $\forall \mathbf{x} \phi^{R\text{-Skol}}(\mathbf{x})$  its relational Skolemization. Let  $S$  be the original signature of  $\phi$ , and  $S^+$  the relation symbols introduced in the relational Skolemization. Furthermore, for each  $k$ -ary  $R^+ \in S^+$  we introduce a new  $(k-1)$ -ary relation  $R^{++}$ . These new symbols will be used to calibrate the weight of models for the reference proposition  $b()$ . Note that the arity of symbols in  $S^+$  is at least 1, and  $R^{++}$ , thus, is well-defined, but may contain relations of arity 0.

The first formula in our knowledge base is

$$a() \wedge \neg \phi^{R\text{-Skol}}(\mathbf{x}) \quad : \quad 0 \tag{11}$$

We now approximately axiomatize the functional nature of the symbols  $R^+ \in S^+$ . The sentence (6) can be directly encoded as a weighted formula:

$$R^+(\mathbf{x}, y) \wedge R^+(\mathbf{x}, y') \wedge y \neq y' \quad : \quad 0 \tag{12}$$

Next, we would like to enforce (7) by means of a weighted formula. However, (7) encodes the essence of the existential quantifiers we are about to eliminate, and, thus, it is not surprising that this is not possible to enforce strictly. However, we can reward models in which the existential quantification of (7) is satisfied via the weighted formulas

$$a() \wedge R^+(\mathbf{x}, y) \quad : \quad w \quad (R^+ \in S^+) \tag{13}$$

where  $w > 1$  is a weight whose exact value is to be defined later.

We now proceed with constraining models of the reference proposition  $b()$ . First,

$$b() \wedge R(\mathbf{x}) \quad : \quad 0 \quad (R \in S) \tag{14}$$

$$b() \wedge R^+(\mathbf{x}, y) : 0 \quad (R^+ \in S^+) \quad (15)$$

In order to allow  $b()$ -models to gain some weight, we use the extra symbols in  $S^{++}$ :

$$b() \wedge R^{++}(\mathbf{x}) : w \quad (R^{++} \in S^{++}) \quad (16)$$

where  $w$  is the same weight as in (13). To limit the possible interpretations of  $b()$ -models, we also stipulate:

$$b() \wedge \neg R^{++}(\mathbf{x}) : 0 \quad (R^{++} \in S^{++}) \quad (17)$$

The extra symbols must have empty interpretations in  $a()$ -models:

$$a() \wedge R^{++}(\mathbf{x}) : 0 \quad (R^{++} \in S^{++}) \quad (18)$$

Finally, we add:

$$\neg(a() \vee b()) : 0 \quad (19)$$

We now determine (approximately)  $w_n^{KB}(a())$  and  $w_n^{KB}(b())$  for the cases  $n \in \text{spec}(\phi)$  and  $n \notin \text{spec}(\phi)$ .

First, consider  $b()$ : for any  $n$ , there exists exactly one interpretation  $I_{b()} \in \mathcal{I}(D_n, S \cup S^+ \cup S^{++} \cup \{a(), b()\})$  in which  $b()$  is true. This is the interpretation in which all relations in  $S \cup S^+$  are empty ((14),(15)), all relations in  $S^{++}$  are maximal (17), and, in consequence of the latter, because of (18),  $a()$  is false.

Assume that  $S^+ = \{R_1^+, \dots, R_m^+\}$ , where  $R_i^+$  has arity  $k_i + 1$ . Then  $R_i^{++} \in S^{++}$  contributes a factor of  $w^{n^{k_i}}$  to  $w_n^{KB}(b())$ , and the total weight is:

$$w_n^{KB}(b()) = w_n^{KB}(I_{b()}) = w^{n^{k_1} + \dots + n^{k_m}} = w^{K(n)}, \quad (20)$$

using for abbreviation  $K(n) := n^{k_1} + \dots + n^{k_m}$ .

We next turn to  $w_n^{KB}(a())$  in the case  $n \in \text{spec}(\phi)$ . Then there exists at least one interpretation  $I \in \mathcal{I}(D_n, S \cup S^+)$ , in which  $\forall \mathbf{x} \phi^{R-Skol}(\mathbf{x})$  is true, and in which the relations from  $S^+$  have a functional interpretation. We can expand this interpretation to an interpretation in  $\mathcal{I}(n, S \cup S^+ \cup S^{++} \cup \{a(), b()\})$  by giving all relations in  $S^{++}$  an empty interpretation, and setting  $a()$  to true and  $b()$  to false. Then  $I$  does not violate any hard constraint in  $KB$ , and collects from (13) a total weight of  $w^{n^{k_1} + \dots + n^{k_m}}$ . Thus

$$w_n^{KB}(a()) \geq w^{K(n)},$$

and therefore, when  $n \in \text{spec}(\phi)$

$$P_n^{KB}(a()) \geq w_n^{KB}(a()) / (w_n^{KB}(a()) + w_n^{KB}(b())) \geq 1/2.$$

Finally, we have to consider  $w_n^{KB}(a())$  in the case  $n \notin \text{spec}(\phi)$ . For any  $I$  with nonzero weight in which  $a()$  is true, because of (11), also  $\forall \mathbf{x} \phi^{R-Skol}(\mathbf{x})$  must be true. This, now, only is possible when some  $R^+ \in S^+$  is not a functional relation, which, because of (12) can only mean that for some  $\mathbf{x}$  there exists no  $y$  with  $R^+(\mathbf{x}, y)$ .

For a given  $I$ , let

$$l(I) := \sum_{i=1}^m |\{\mathbf{d} \in D^{k_i} \mid \exists e I \models R_i^+(\mathbf{d}, e)\}| \quad (21)$$

In the case where  $k_i = 0$ , the summation term on the right of (21) is 1 if  $R_i^+(e)$  holds for some  $e$ , and 0 otherwise. Thus,  $l(I)$  counts the number of  $\mathbf{d}$  that substituted for  $\mathbf{x}$  in (13) contribute a weight factor  $w$  to  $w(I)$ . The weight of  $I$  with  $I \models a()$  then is  $w^{l(I)}$  (noting that because of (18)  $I$  cannot obtain any additional weight from interpretations of  $R^{++}$  relations).

We now count the number of interpretations with the precise weight  $w^l$  collected from (13). There are  $n^{K(n)}$  different  $\mathbf{d}$ ,  $R^+$ -combinations  $R^+(\mathbf{d}, \cdot)$  that can lead to an increment of 1 to the sum (21). Thus, there are  $\binom{n^{K(n)}}{l}$  different selections of  $R^+(\mathbf{d}, \cdot)$  to obtain a sum of  $l$ . For each such selection, there are  $n^l$  different choices for  $e$  to make  $R^+(\mathbf{d}, e)$  true. Thus, there is a total of  $\binom{n^{K(n)}}{l} n^l$  different interpretations of the relations in  $S^+$  to obtain a weight of  $w^l$ . For each such interpretation in which  $a()$  also is true, the relations of  $S^{++}$  are empty, and  $b()$  is false. However, we still have to take into account possible interpretations of the relations in  $S$ . If  $L(n)$  is the total number of ground  $S$ -atoms  $R(\mathbf{d})$  ( $R \in S$ ), then there are  $2^{L(n)}$  different interpretations for  $S$ .  $L(n)$ , like  $K(n)$ , is a polynomial in  $n$ . Thus, we can bound

$$w_n^{KB}(a()) \leq \sum_{l=0}^{K(n)-1} \binom{n^{K(n)}}{l} n^l 2^{L(n)} w^l \quad (22)$$

where the sum is over all possible values of  $l$  in which at least one symbol in  $R^+$  does not have a functional interpretation. Note that the right side of (22) may give a rather extreme over-estimate of  $w_n^{KB}(a())$ , since most of the interpretations of  $S \cup S^+$  that are counted here with a weight of  $w^l$  may make  $\neg \phi^{R-Skol}(\mathbf{d})$  true for some  $\mathbf{d}$ , and, thus, by (11) have an actual weight of 0.

By further lower-bounding the right side of (22), we obtain

$$\begin{aligned} w_n^{KB}(a()) &\leq \sum_{l=0}^{K(n)-1} n^{lK(n)} n^l 2^{L(n)} w^l \\ &\leq K(n) n^{(K(n)-1)K(n)} n^{K(n)-1} \cdot 2^{L(n)} w^{K(n)-1} \\ &\leq n^{M(n)} w^{K(n)-1} \end{aligned}$$

where  $M(n) = (K(n)-1)K(n) + (K(n)-1) + L(n)$  is a polynomial in  $n$ . We now obtain for the case  $n \notin \text{spec}(\phi)$

$$\begin{aligned} P_n^{KB}(a()) &\leq w_n^{KB}(a()) / w_n^{KB}(b()) \\ &\leq n^{M(n)} w^{(K(n)-1)} / w^{K(n)} = n^{M(n)} / w. \end{aligned}$$

Setting  $w = 10n^{M(n)}$ , we thus have  $P_n^{KB}(a()) \leq 1/10$  if  $n \notin \text{spec}(\phi)$ . The representation size of  $w$  is polynomial in  $n$ . Thus, an algorithm that computes  $P_n^{KB}(a())$  up

to an accuracy of  $0.2 = (0.5 - 0.1)/2$  in time polynomial in  $n$  and the representation size of  $w$  would give a polynomial time decision procedure for  $\text{spec}(\phi)$ .  $\square$

## References

- [1] J. S. Breese. Construction of belief and decision networks. *Computational Intelligence*, 8(4):624–647, 1992.
- [2] R. de Salvo Braz, E. Amir, and D. Roth. Lifted first-order probabilistic inference. In *Proceedings of the Nineteenth Int. Joint Conf. on Artificial Intelligence (IJCAI-05)*, pages 1319–1325, 2005.
- [3] R. de Salvo Braz, S. Natarajan, H. Bui, J. Shavlik, and S. Russell. Anytime lifted belief propagation. In *Proc. of SRL 2009*, 2009.
- [4] Ronald Fagin. Probabilities on finite models. *Journal of Symbolic Logic*, 41(1):50–58, 1976.
- [5] D. Fierens, G. Van den Broeck, I. Thon, B. Gutmann, and L. De Raedt. Inference in probabilistic logic programs using weighted cnf’s. In *Proc. of UAI 2011*, 2011.
- [6] N. Friedman, Lise Getoor, D. Koller, and A. Pfeffer. Learning probabilistic relational models. In *Proceedings of the 16th International Joint Conference on Artificial Intelligence (IJCAI-99)*, 1999.
- [7] V. Gogate and P. Domingos. Probabilistic theorem proving. In *Proc. of UAI 2011*, 2011.
- [8] M. Jaeger. On the complexity of inference about probabilistic relational models. *Artificial Intelligence*, 117:297–308, 2000.
- [9] M.\* Jaeger. Relational bayesian networks. In Dan Geiger and Prakash Pundalik Shenoy, editors, *Proceedings of the 13th Conference of Uncertainty in Artificial Intelligence (UAI-13)*, pages 266–273, Providence, USA, 1997. Morgan Kaufmann.
- [10] A. Jha, V. Gogate, A. Meliou, and D. Suciu. Lifted inference seen from the other side: The tractable features. In *Proc. of NIPS*, 2010.
- [11] D. S. Johnson. A catalog of complexity classes. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science*, volume 1, pages 67–161. Elsevier, Amsterdam, 1990.
- [12] N. D. Jones and A. L. Selman. Turing machines and the spectra of first-order formulas with equality. In *Proceedings of the Fourth ACM Symposium on Theory of Computing*, pages 157–167, 1972.

- [13] K. Kersting and L. De Raedt. Towards combining inductive logic programming with bayesian networks. In *Proceedings of the 11th International Conference on Inductive Logic Programming (ILP-01)*, volume 2157 of *LNAI*, pages 118–131, 2001.
- [14] J. Kisiński and D. Poole. Lifted aggregation in directed first-order probabilistic models. In *Proc. of IJCAI 2009*, 2009.
- [15] B. Milch, B. Marthi, S. Russell, D. Sontag, D.L. Ong, and A. Kolobov. Blog: Probabilistic logic with unknown objects. In *Proc. 19th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1352–1359, 2005.
- [16] B. Milch, L. S. Zettlemoyer, K. Kersting, M. Haimes, and L. P. Kaelbling. Lifted probabilistic inference with counting formulas. In *Proc. of AAAI 2008*, 2008.
- [17] L. Ngo, P. Haddawy, and J. Helwig. A theoretical framework for context-sensitive temporal probability model construction with application to plan projection. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, pages 419–426, 1995.
- [18] D. Poole. Probabilistic horn abduction and Bayesian networks. *Artificial Intelligence*, 64:81–129, 1993.
- [19] D. Poole. First-order probabilistic inference. In *Proceedings of IJCAI-03*, 2003.
- [20] M. Richardson and P. Domingos. Markov logic networks. *Machine Learning*, 62(1-2):107 – 136, 2006.
- [21] T. Sato. A statistical learning method for logic programs with distribution semantics. In *Proceedings of the 12th International Conference on Logic Programming (ICLP’95)*, pages 715–729, 1995.
- [22] P. Sen, A. Deshpande, and L. Getoor. Bisimulation-based approximate lifted inference. In *Proc. of UAI*, 2009.
- [23] P. Singla, A. Nath, and P. Domingos. Approximate lifted belief propagation. In *Proc. of AAAI-10 Workshop on Statistical Relational AI*, 2010.
- [24] B. Taskar, P. Abbeel, and D. Koller. Discriminative probabilistic models for relational data. In *Proc. of UAI 2002*, 2002.
- [25] G. Van den Broeck. On the completeness of first-order knowledge compilation for lifted probabilistic inference. In *Proc. of the 25th Annual Conf. on Neural Information Processing Systems (NIPS)*, 2011.
- [26] G. Van den Broeck, N. Taghipour, W. Meert, J. Davis, and L. De Raedt. Lifted probabilistic inference by first-order knowledge compilation. In *Proc. of IJCAI 2011*, 2011.

- [27] Joost Vennekens, Marc Denecker, and Maurice Bruynooghe. Representing causal information about a probabilistic process. In *Logics in Artificial Intelligence, 10th European Conference, JELIA 2006, Proceedings*, volume 4160 of *Lecture Notes in Computer Science*, pages 452–464. Springer, 2006.